查询专指度对检索效果的影响研究

任　珂[1]　陆　伟[1,2]　丁　恒[1]
[1](武汉大学信息管理学院　武汉　430072)
[2](武汉大学信息资源研究中心　武汉　430072)

摘要:【目的】针对不同查询专指度语句的检索效果进行全面分析, 为改善搜索引擎性能、提高用户检索体验提供借鉴。【方法】基于 TREC Web Track 查询语句, 人工构建查询专指度标注集, 选用语言模型狄利克雷平滑、语言模型线性插值平滑和 BM25 三种模型, 以常用的信息检索评价指标为基准, 探讨查询专指度强弱对检索效果在不同层次上的影响。【结果】在最靠前的几条检索结果中, 强弱专指度查询语句的检索效果差异最大, 强专指度的检索效果要明显好于弱专指度。【局限】仅在 TREC 数据集上进行实验测试, 还需在其他数据集上进一步检验。【结论】搜索引擎在专指度这一维度下, 应重点关注最靠前的几条检索结果的准确性, 以此为切入点改善检索模型。

关键词: 查询意图　查询专指度　检索效果
分类号: G353.1

1　引　言

互联网为人们提供了内容丰富、形式多样的信息资源, 信息的丰富性和多样性不仅给人们带来了便利, 同时也导致了信息过载[1], 并使得用户信息搜寻的难度增大。Google、百度等搜索引擎作为用户与网络资源交互的接口, 会根据用户输入的查询返回结果列表, 用户输入的查询过于宽泛或过于局限都有可能造成检索结果出现偏差, 只能通过手动筛选检索结果或二次查询进行检索结果的修正和获取, 造成用户时间和精力的浪费, 在一定程度上造成信息的遗失。因此针对不同查询专指度(Query Specificity)的用户查询语句, 制定不同的检索策略, 是检索模型改进的方向之一。

查询专指度属于查询语句的语义特征[2], 是影响信息检索效果的因素之一[3], 查询专指度反映了查询语句表达概念的宽泛性, 集中体现用户对所检索信息的详细性和确定性要求[4], 在一定程度上表明用户查询意图。加强对查询语句专指度的研究可以帮助搜索引擎更好地理解用户的潜在需求, 提供更符合其信息需求的检索结果列表。

目前, 学界对于“查询专指度强弱程度对检索效果的影响”这一问题只关注了查询专指度对整个检索列表的影响, 但由于不同用户的搜索、点击行为具有个体差异性, 以整个检索列表为研究对象不利于深入了解用户的多样性需求。本文根据查询专指度的定义, 梳理归纳了查询专指度的分类依据, 对 Text Retrieval Conference (TREC) Web Track 2009 年–2012 年发布的所有查询语句进行人工分类, 并以常用的信息检索评价指标为基准, 全方位分析查询专指度对检索效果的影响, 以便在这一维度对检索效果进行改善。

2　相关研究

查询意图是介于用户查询语句与用户真实信息需求之间的一种中间形式, 用于表示用户的搜索目的[5], 查询意图的深层次分析有利于构造用户的信息需求空间, 明确用户搜索意图以及据此提供更直接、更相关和更丰富的信息。2002 年, Broder[6]从用户搜索目的角度出发将查询分为导航类、信息类和事务类三大类, 随后有学者在此基础上进一步细化和调整[7-8], 但基本延续其框架体系。但由于这一分类体系过于简单, 难

通讯作者: 陆伟, ORCID: 0000-0002-0929-7416, E-mail: reedwhu@gmail.com。

以应对具有复杂信息需求的查询，因此 González-Caro 等[9]提出从题材(Genre)、主题(Topic)、专指度(Specificity)、任务(Task)、目标(Objective)、范围(Scope)等 10 大维度揭示用户复杂的信息需求和丰富的信息空间，本文即以专指度为立足点进行用户查询意图分析。

查询专指度是搜索引擎查询的一个重要方面，但由于专指度的语义特性致使其很难测量，目前针对查询专指度的研究主要可以分为三个方面：

(1) 查询专指度的分类。这个方面主要是针对专指度这一维度进行深入研究，了解其内涵和特征，主要分为两种方式，一种是按查询专指度的强弱程度分类，比如 Hafernik[10]利用 9 个语义属性将查询专指度分为强弱两类，同时发现查询长度和词性可以提高专指度强弱识别的准确性，唐祥彬等[4]对查询语句的特征进行分析，并利用机器学习算法自动识别查询语句的专指度强弱；另一种是利用专指度的不同表示方式将其分为检索文档数量代表的专指度、词形变换数量代表的专指度和专业领域词汇数量代表的专指度三类，并判断这三类之间的相关性[11]。

(2) 查询专指度和其他属性的关系。这个方面主要是将专指度放入检索属性网络中，研究各个属性之间的关系，增强检索网络的一致性和完整性。Phan 等[12]发现查询专指度与查询语句长度相关，长度越短则专指度越小，且三个词的长度通常为强弱专指度的分界线；Kim[3]则着眼于查询专指度与文档相关性之间的关系，回归查询本质。

(3) 查询专指度对检索效果的影响。学界研究查询专指度的目的是为增强检索效果，Mu 等[13]将查询专指度和查询语句长度作为查询扩展的两种方式，研究在健康信息中增大或减少专指度和语句长度对整体检索效果的影响；Heine [14]研究数据库中信息量、查询语句长度和平均查询专指度三个因素对 MEDLINE 数据库中信息检索效果的影响，且发现在此数据库中专指度影响并不突出。

以上研究表明，有关查询专指度的研究还存在以下不足：

(1) “查询专指度对检索效果影响”的研究集中在医学信息检索任务，没有对互联网开放环境下的检索情境进行分析；

(2) 研究仅仅探讨了查询专指度与整体检索效果的关系，没有在不同指标下对检索结果进行不同层次的研究。

针对这些不足，本文在借鉴前人研究基础上完善查询专指度分类体系特征，并在 TREC Web Track 数据集上构建查询专指度强弱程度标注集，利用 BM25 模型和语言模型等检索模型，以多种常用的信息检索评价指标为基准对查询语句的检索效果进行全面分析，同时探讨检索模型在不同专指度下的检索效果。

3 查询专指度强弱程度分类及特征选择

3.1 查询专指度强弱程度分类

不同专指度强弱在不同语境中可能会产生不同的检索效果，通过对数据集中的查询语句进行分类，分别比较不同强弱程度下的检索效果是本研究的目的。目前，学界对查询专指度强弱程度没有明确的分类标准，一般将其分为两类或三类，两类为强专指度(Narrow)和弱专指度(Broad)，三类为强专指度(Specific)、略专指度(Medium)和弱专指度(Broad)。通常地，由于英文查询语句长度较短，词性短小精悍，因此本文从信息需求角度出发[15-16]，将查询专指度分为强弱两类。

(1) 强专指度查询：用户表达出明确的信息需求，且此需求产生的歧义较少或不产生歧义，清晰地表达了用户的目的和查询需求的范围，或涉及到某些专业领域知识。比如用户想知道某一问题的确切答案、对某一观点或话题进行比较、想知道某一确切日期发生的事情等，如查询“who invented music”、“mothers day songs”等。

(2) 弱专指度查询：用户表达的信息需求不明确，由此产生较大歧义，或用户表达的查询目的和需求范围较广，属于一般领域，无法进行准确定位。这类查询往往要进行二次检索或人工筛选，如“cell phones”、“korean language”和“dieting”等。

3.2 查询专指度强弱程度特征分析

查询语句特征分为基本特征和内容特征，基本特征为查询语句的长度和词项个数等，内容特征关注查询语句的含义。查询专指度作为查询语句的语义特征，主要针对查询专指度的内容特征进行分析。专指度通常关注在用户查询中使用了哪些限制来明确用户需求，如数量限制、名字限制、时间限制、位置限制等，基

于此，并结合 Hafernik[10]和唐祥彬等[4]的研究，本文针对专指度强弱程度分类选择 9 个属性特征建立属性列表，如表 1 所示：

表 1　查询语句属性特征及查询举例

编号	查询语句属性特征	查询举例
1	查询语句包括 URL 或网站名称或 IP 地址	yahoo
2	查询语句包含确切地方名称及其他词项	map of the united states
3	查询语句比较不同事物或同一事物的不同方面	butter and margarine
4	查询语句比较多种不同的观点或话题	keyboard reviews
5	查询语句为包含确切答案的问题	who invented music
6	查询语句包含方向、建议、指导等方面的信息需求	how to build a fence
7	查询语句包含确切时间及其他词项	mothers day songs
8	查询语句包含特定数字及其他词项	HP mini 2140
9	查询语句包含事物名称及其他词项	Obama family tree

由于查询语句所代表的用户查询意图对于理解专指度有重要的意义，因此所选择的属性一方面表明用户的查询需求，另一方面是基于前人研究选择可以表征强专指度的属性特征，例如如果一个查询语句包含一个网址，则表明用户希望找到某一特定网站，属于强专指度查询，又如一个查询语句包含对于某种事情的指导等信息需求，则表明用户希望寻找达到这一目标的一系列具体步骤。本文基于表 1 中的查询语句的属性特征，将查询专指度分为强弱两类，当包含上述一个或多个属性特征时即为强专指度查询，不包含任一属性特征时即为弱专指度查询。

4　研究对象及研究思路

4.1　研究对象

本文数据集来自 TREC Web Track 2009 年–2012 年发布的查询语句，每年发布 50 个，总共 200 个查询语句，由于这些查询语句均为网页数据，主要针对互联网开放环境下的检索情景，即是普通大众最常接触的信息搜索情景，因此可以最大限度地保证研究结果的实用性。在此数据集上，利用 Indri5.7 建立索引，并利用 Indri 标准停用词表进行停用词过滤，同时利用 Krovetz 进行词干提取。根据表 1 中建立的查询语句属性特征对其进行专指度强弱程度标注，标注工作由两名武汉大学信息管理学院图情专业硕士研究生完成，为检验个人主观因素对分类标注的干扰程度，检验其标注 Kappa 统计量[17](Kappa Statistic)，实验表明 Kappa 值为 0.91，大于 0.8，说明实验标注结果存在很好的一致性，具有研究意义。对于当前大规模语料集合，在对每个查询语句的检索结果进行相关性评价时，列举每个查询的所有相关文档是不现实的，因此通常利用缓冲池法[18](Pooling)，即将一系列检索系统中每个系统所返回的前 K 篇文档合成一个文档子集，并对这个子集进行相关性判定，本文利用 TREC 的评价基准针对每个查询语句的 Top1000 文档进行测评。在所有 200 个查询语句中，有两个查询语句在 Track 提供的相关性文档集和基准评价中是不存在的，因此最终只针对 198 个查询语句进行检索结果分析。

4.2　研究方法

在各种信息检索模型中，应用最广泛的为语言模型[19](Language Model)和 Okapi BM25 模型[20](Okapi BM25 Model)，因此本文运用这两种模型对查询语句进行检索排序，以减少单一模型可能出现的检索偏差。同时为了避免词项在文档中出现“零概率”问题，在语言模型中使用两种平滑方式，分别为狄利克雷平滑(Dirichlet)和线性插值平滑(Jelinek-Mercer)。在检索模型基础之上，利用 trec_eval 计算不同模型下的评价指标，进而评价其检索结果的差异。评价查询语句的检索效果通常基于传统的召回率(Recall)和准确率(Precision)测量方法，但这两种方法都是基于集合的评价方法，在面对搜索引擎等系统输出的有序检索结果时，针对性较差，因此运用平均正确率均值 (Mean Average Precision, MAP)、归一化折损累积增益(Normalized Discounted Cumulative Gain, NDCG)、R 正确率(R-Prec)、Bpref (Binary preference-based measures)、Recip_Rank、P@5、P@10 和 P@20 等检索领域常用的用于评价有序检索结果的指标，对检索结果进行更加全面和准确的评价。

(1) MAP：反映检索系统在全部相关文档上性能的单值指标。

(2) NDCG：衡量检索结果中全部文档排序质量的指标。

(3) R-Prec：检索结果中前 R 个结果集的正确率，R 是当前检索中相关文档总数。

(4) Bpref：重点关注不相关文档在相关文档之前出现的次数。

(5) Recip_Rank：表示检索系统返回第一个相关文档的能力。

(6) P@K：反映检索系统对于查询语句返回的前 K 个检索结果的准确率。

从上述各个检索指标的含义可以看出，不同指标有不同的意义，MAP、NDCG 等指标是从整体检索效果上进行评价的，而 Recip_Rank、P@K 等指标是对结果列表相对靠前的检索结果进行评价，通过运用不同的指标，可以反映查询专指度对检索结果在不同层次上的影响作用。

4.3　研究框架

通过对 198 个查询语句按照专指度强弱进行人工分类，并在语言模型狄利克雷平滑、语言模型线性插值平滑和 BM25 三个模型中分别对所有查询语句的 Top1000 的检索结果文档进行相关性评价，得到包括相关文档数量及 MAP、NDCG、R-Prec、Bpref、Recip_Rank、P@5、P@10 和 P@20 等各指标值，并在此基础上，分别以"同一模型"和"同一专指度强弱"为基准点，以不同指标为评价方式，对检索效果在专指度这一维度下进行全面分析，具体来说，"同一模型下专指度强弱对检索效果的影响"重点关注不同专指度查询语句检索效果的差异性，"同一专指度强弱下不同模型的比较"重点关注针对同一专指度强弱，不同模型间的性能差异，通过这两方面的研究，可以立体、全面地描述查询专指度对检索效果的影响，为改善搜索引擎性能、提高用户检索体验提供思路。

5　专指度强弱对检索效果的影响

运用统计学指标(均值、标准差、中值、最小值、最大值)对专指度强弱对检索效果的影响及检索模型的比较进行描述，同时运用箱形图对其进行可视化展示，利用曼–惠特尼秩和检验(Mann-Whitney U Test)对两样本进行显著性差异判断，若曼–惠特尼秩和检验值小于 0.05，则拒绝原无差异假设，认为两者有统计学差异。

5.1　同一模型下对强弱专指度的检索比较

分别在语言模型狄利克雷平滑、语言模型线性插值平滑和 BM25 三种模型下，运用 MAP、NDCG、R-Prec、Bpref、Recip_Rank、P@5、P@10 和 P@20 等指标，对不同专指度的查询语句检索效果进行对比。表 2 为在三种检索模型下针对特定指标不同专指度查询语句的检索结果的曼–惠特尼秩和检验值。

表 2　不同检索模型下评价指标的对比

模型	MAP	NDCG	R-Pref	Bpref	Recip_Rank	P@5	P@10	P@20
语言模型狄利克雷平滑	0.715	0.594	0.763	0.497	0.095	**0.036**	0.108	0.454
语言模型线性插值平滑	**0.04**	0.115	0.216	0.452	**0.029**	**0.047**	0.111	0.068
BM25 模型	0.23	0.74	0.26	0.426	**0.012**	**0.012**	0.105	0.078

从表 2 中可以看出，在语言模型狄利克雷平滑下 P@5 指标值为 0.036，小于 0.05，表现出强弱专指度检索效果的差异；在语言模型线性插值平滑下 MAP、Recip_Rank 和 P@5 指标值分别为 0.04、0.029 和 0.047，也均低于 0.05，具有检索效果的统计学差异；而在 BM25 模型下，Recipe_Rank 和 P@5 指标也表现出同样的统计学差异。同时，这些具有统计学差异的指标值(6 组)占到全部统计指标值(24 组)的 25%，且在每一种模型下都有指标显示查询专指度的不同强弱程度会造成检索结果的不同。以上 6 组具有统计学差异的箱形图如图 1 所示。

具体来说，在三种模型下 P@5 都存在统计学差异，这说明在不同专指度强弱查询语句下，检索列表中前 5 条检索结果的准确率是明显不同的，从图 1(a)，(d)，(f)中可以看出，三种模型强专指度查询下 P@5 的最大值、平均值等统计学指标均大于弱专指度，且极少出现离群点，说明在 P@5 指标下强专指度的检索效果要明显好于弱专指度，这是由于弱专指度查询语句查询范围较广，一般不对查询语句进行限制，因此最先出现的结果会包含事物的不同方面，比如查询"apple"，既有可能出现苹果公司的网站信息，也有可能出现 "苹果"这种水果的相关信息，造成与用户查询

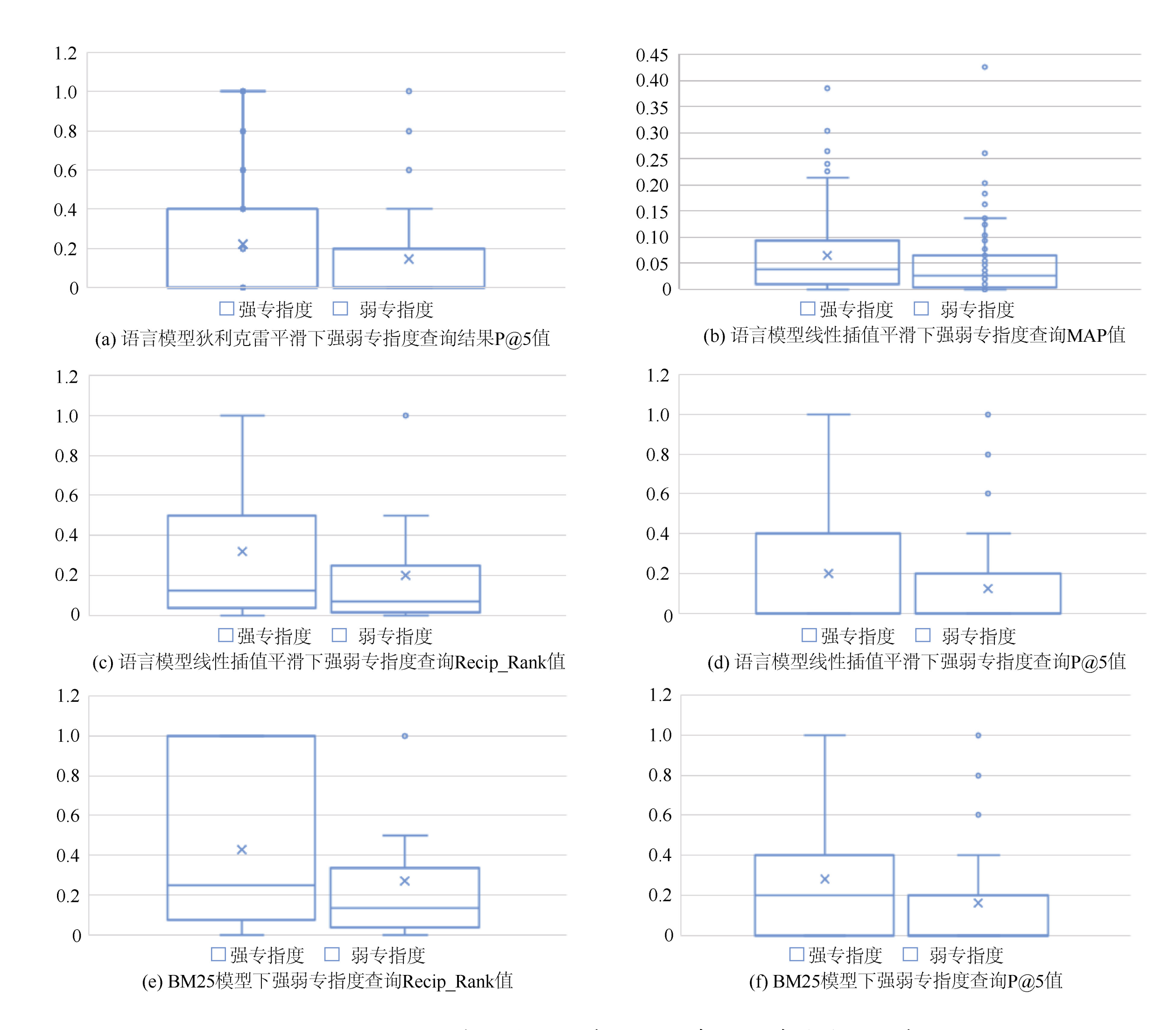


图 1 同一模型下强弱专指度的检索效果具有统计学差异的检索指标值

意图背离的结果，而强专指度查询由于对用户需求进行了一定的限制，歧义较小，因此在检索结果的前 5 条中强专指度的检索效果要明显好于弱专指度。

除此之外出现较多差异的指标是 Recip_Rank，在语言模型线性插值平滑和 BM25 两种模型下强弱专指度查询的曼–惠特尼秩和检验值分别为 0.029 和 0.012，小于 0.05，具有统计学差异。Recip_Rank 作为表示返回第一个相关结果能力的指标，从一定程度上表征了第一个返回结果的相关性强弱，在图 1(c), (e)中可以看出，在此指标下，强专指度的效果要明显好于弱专指度，在最大值和平均值方面都有较大的优势，原因主要为强专指度的查询一般都会有比较明确的答案，或是话题或想法相对集中，因此在进行检索的过程中通常会最先出现用户常用的结果。相较于 P@5 指标下的检索结果，Recip_Rank 表现出强弱专指度查询检索结果的更大的不稳定性。

最后在语言模型线性插值平滑下强弱专指度的 MAP 值也有一定的差异性，曼–惠特尼秩和检验值为 0.04，接近 0.05，由于 MAP 主要针对所有相关文档的准确率进行平均求值，因此强弱专指度检索效果的差异性也逐渐减弱，从图 1(b)中也可以看出，两者差异较小，只有在离群点方面具有明显不同。

通过不同指标可以看出，在涉及到较少的检索结果时，强专指度查询语句的检索效果要明显好于弱专指度，随着检索结果的不断增加，这种差异性逐渐减弱，但总体而言即使在没有显著差异的指标下强专指

度检索结果的平均值也要略好于弱专指度。而从检索模型角度看，语言模型狄利克雷平滑对于强弱专指度的检索效果相对其他两种较好，除在 P@5 指标下有差异外，其他结果的差异性均不大，表明对强弱两种专指度的识别和检索具有一定的适应性和敏感性，是比较综合的检索模型，而其他两种模型则对弱专指度的检索效果较差。

5.2　同一专指度下不同模型的比较

在同一专指度下对三种模型在不同评价指标下的统计学指标(均值、标准差、中值、最小值、最大值)进行比较，采用两两比较的方式，充分展示面对同一强弱专指度查询时，不同检索模型的表现情况。表 3 至表 5 是三种模型两两比较时不同评价指标下曼–惠特尼秩和检验值。

表 3　语言模型狄利克雷平滑与语言模型线性插值平滑比较

比较项	MAP	NDCG	R-Pref	Bpref	Recip_Rank	P@5	P@10	P@20
强专指度查询	**0.01**	**0.016**	**0.008**	0.235	0.633	0.583	0.066	**0.029**
弱专指度查询	**0**	**0**	**0.002**	0.177	0.176	0.736	0.087	**0.003**

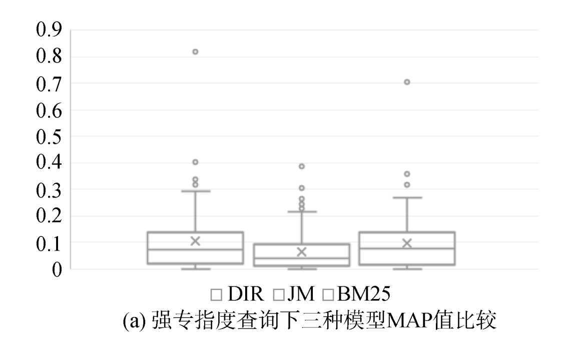
表 4　语言模型狄利克雷平滑与 BM25 模型比较

比较项	MAP	NDCG	R-Pref	Bpref	Recip_Rank	P@5	P@10	P@20
强专指度查询	0.856	0.751	0.841	0.928	0.085	0.192	0.652	0.686
弱专指度查询	0.327	0.224	0.359	0.844	0.439	0.305	0.662	0.566

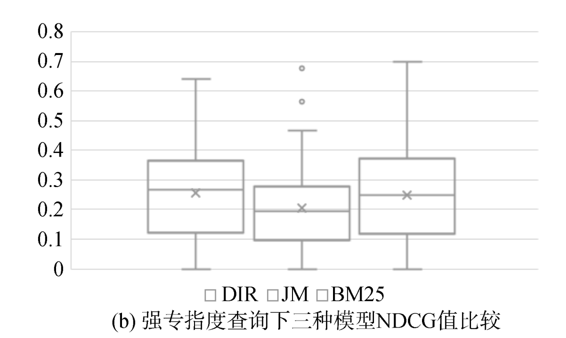
表 5　语言模型线性插值平滑与 BM25 模型的比较

比较项	MAP	NDCG	R-Pref	Bpref	Recip_Rank	P@5	P@10	P@20
强专指度查询	**0.014**	**0.037**	**0.018**	0.257	**0.032**	0.065	**0.025**	**0.006**
弱专指度查询	**0.002**	**0.002**	**0.018**	0.254	**0.042**	0.156	**0.026**	**0.011**

从表 3 可以看出，在语言模型狄利克雷平滑与语言模型线性插值平滑的对比中，强弱专指度下 MAP、NDCG、R-Pref 和 P@20 的曼–惠特尼秩和检验值均小于 0.05，具有统计学差异；而在语言模型狄利克雷平滑与BM25模型比较时，所有指标的曼–惠特尼秩和检验值均大于 0.05，不具有统计学差异，说明这两种模型在同一专指度查询下性能基本一致；而在语言模型线性插值平滑与 BM25 的比较中，出现统计学差异的指标增多，分别为 MAP、NDCG、R-Pref、Recip_Rank、P@10 和 P@20，在强弱专指度查询语句下都小于 0.05。总体来看，在两两模型比较下，强弱专指度查询结果具有的一定的同步性，即当强专指度下模型具有一定的统计学差异时，弱专指度也具有统计学差异。具有统计学差异的模型比较箱形图如图 2 所示:



(a) 强专指度查询下三种模型MAP值比较



(b) 强专指度查询下三种模型NDCG值比较

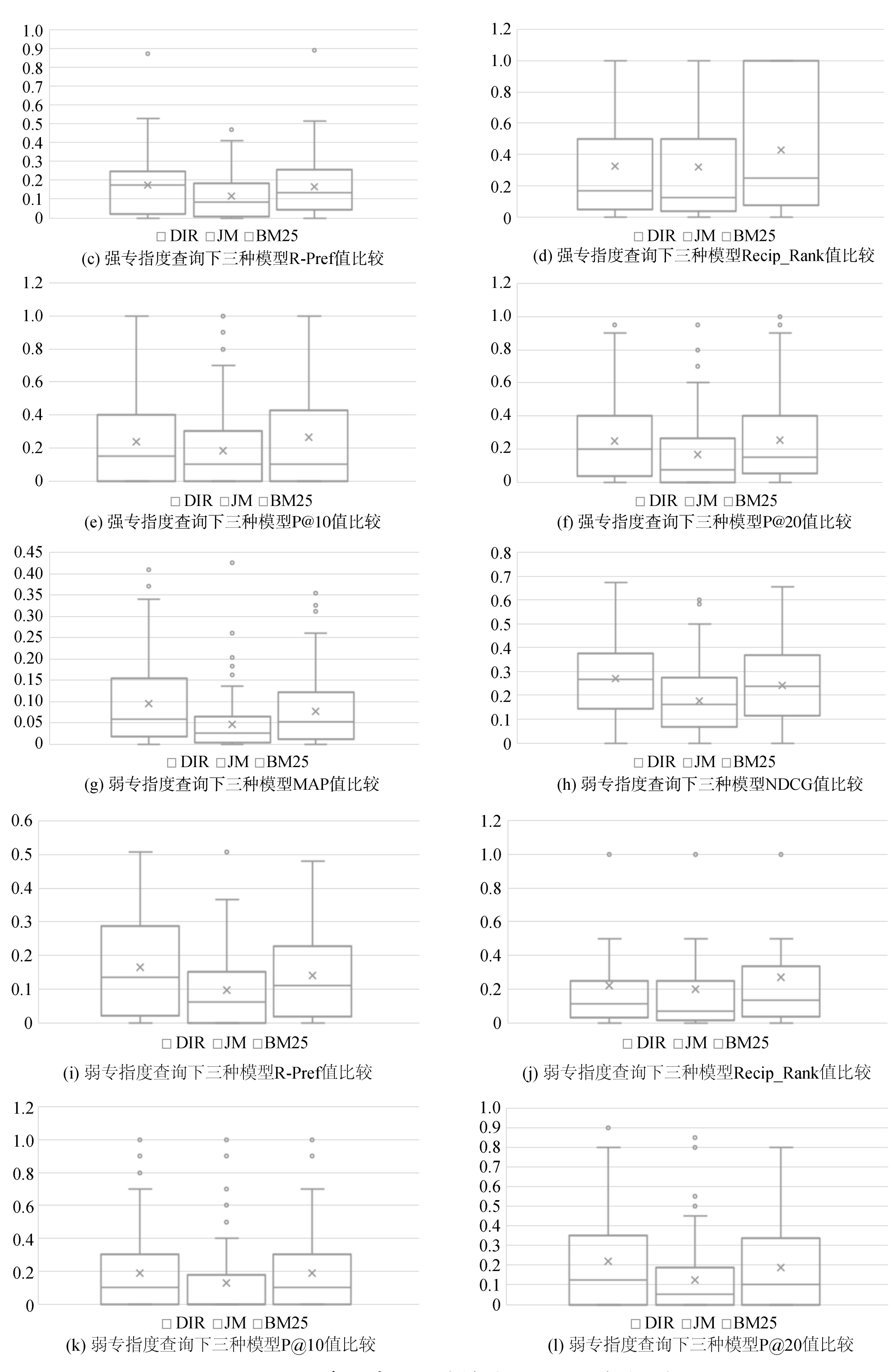


图 2　同一专指度下三种模型不同指标值的比较

从图 2 和表 4 可以看出，语言模型狄利克雷平滑与 BM25 模型在不同指标下的最大值、最小值和平均值等大体相同，曼–惠特尼秩和检验在各个指标下也均大于 0.05，说明这两个模型在针对同一强弱专指度查询语句时，检索效果基本一致，不具有差异性；同时，语言模型狄利克雷平滑和 BM25 模型在箱形图的四分位数和中位数等数值上均比语言模型线性插值平滑高，因此三种模型对于同一强弱专指度的查询语句进行检索时，检索效果由好到差依次为：语言模型狄利克雷平滑=BM25>语言模型线性插值平滑。因此在针对专指度这一维度进行检索时，运用语言模型狄利克雷平滑和 BM25 两种模型会得到更符合用户需求的检索结果。

6 讨 论

从上面的分析中可以看出，在不同检索评价指标下，由于强弱专指度查询在语句明确性方面的差异，在判断最靠前的几条检索结果时，两者的检索效果差距较大，强专指度查询语句的检索效果要明显好于弱专指度，而当对所有检索结果的相关性进行平均判断时，检索效果差异性减小，也就是说，当返回越多的检索文档时，强弱专指度查询语句检索效果的差异性就会减弱，使两者呈现基本一致状态。同时用户在进行检索时，通常只会查看前 10 条或前 20 条检索结果，因此，在专指度这一维度下，检索模型应重点关注如何提高最靠前的几条检索结果的准确率，以实现模型优化。另一方面，在针对同一查询专指度进行检索时，语言模型狄利克雷平滑与 BM25 模型的检索效果要明显好于语言模型线性插值平滑。因此，搜索引擎可以根据查询专指度强弱程度特征，利用机器学习算法对专指度进行自动识别，同时搜索引擎可以充分利用语言模型狄利克雷平滑或 BM25 模型的检索优势，返回与用户信息需求限制范围相符的个性化查询结果，以此改善搜索引擎性能，提高用户体验。

除此之外，通过实验发现：

(1) 在所有的检索指标中，Bpref 是唯一一个在同一模型的强弱专指度比较和同一强弱专指度下不同模型比较两个方式中均不存在统计学差异的指标，且数值均较大，分析认为这是由于 Bpref 指标重点关注不相关文档在相关文档之前出现的次数，而对于 198 个 TREC 查询语句，在每个查询返回的 1 000 篇文档中，相关文档数量较少，最大为 167 篇，平均为 37 篇，数量上处于中等偏下的程度，因此在此指标下差别不明显，显示不出统计学差异，但并不代表模型在此方面没有缺陷，未来可进一步研究。

(2) 对于 P@5 指标来说，它在同一模型下对强弱专指度检索结果进行评价时具有统计学差异，而在针对同一专指度强弱比较不同模型时却没有统计学差异，说明在检索结果的前 5 条(也就是最重要的 5 条中)各个模型都对强专指度有着更好的检索结果，而对于弱专指度的检索就比较差强人意，需要进一步改进。

7 结 语

根据不同查询专指度，返回与其信息需求限制范围相符的个性化查询结果，成为改善搜索引擎性能、提高用户检索体验的重要途径。本文基于 TREC Web Track 查询语句，人工构建查询专指度标注集，选用语言模型狄利克雷平滑、语言模型线性插值平滑和 BM25 三种模型，以常用的信息检索评价指标为基准，对检索效果在专指度这一维度下进行了全面分析，实验结果表明：在最靠前的几条检索结果中，强弱专指度查询语句的检索效果差异最大，强专指度的检索效果要明显好于弱专指度。同时本文研究也有一定的局限性，仅在 TREC 数据集上进行实验测试，还需在其他数据集上进一步检验。

参考文献：

[1] 王娜，陈会敏．泛在网络中信息过载危害及原因的调查分析[J]．情报理论与实践，2014，37(11)：20-25．(Wang Na, Chen Huimin. Investigation on the Harm and Cause of Information Overload in Ubiquitous Network [J]. Information Studies: Theory & Application, 2014, 37(11): 20-25.)

[2] Jones K S. A Statistical Interpretation of Term Specificity and Its Application in Retrieval [J]. Journal of Documentation, 1972, 28(1): 11-21.

[3] Kim G. Relationship Between Index Term Specificity and Relevance Judgment [J]. Information Processing & Management, 2006, 42(5): 1218-1229.

[4] 唐祥彬，陆伟，张晓娟，等．查询专指度特征分析与自动识别[J]．现代图书情报技术，2015(2)：15-23．(Tang Xiangbin, Lu Wei, Zhang Xiaojuan, et al. Feature Analysis

and Automatic Identification of Query Specificity [J]. New Technology of Library and Information Service, 2015(2): 15-23.)

[5] 宋巍. 基于主题的查询意图识别研究[D].哈尔滨: 哈尔滨工业大学, 2013. (Song Wei. Research on Topic Based Query Intent Identification [D]. Harbin: Harbin Institute of Technology, 2013.)

[6] Broder A. A Taxonomy of Web Search [J]. ACM SIGIR Forum, 2002, 36(2): 3-10.

[7] Rose D E, Levinson D. Understanding User Goals in Web Search [C]. In: Proceedings of the 13th International Conference on World Wide Web. New York, NY, USA: ACM, 2004: 13-19.

[8] Baeza-Yates R, Calderón-Benavides L, González-Caro C. The Intention Behind Web Queries [C]. In: Proceedings of the 13th International Conference on String Processing and Information Retrieval. Berlin, Heidelberg: Springer-Verlag, 2006: 98-109.

[9] González-Caro C, Calderón-Benavides L, Baeza-Yates R, et al. Web Queries: The Tip of the Iceberg of the User's Intent [C]. In: Proceedings of the 4th ACM WSDM Conference, Hong Kong, China. 2011.

[10] Hafernik C T. The Relationship Between Query Length, Parts of Speech Usage and Web Search Query Specificity [D]. The Pennsylvania State University, 2013.

[11] Tamine L, Chouquet C, Palmer T. Analysis of Biomedical and Health Queries: Lessons Learned from TREC and CLEF Evaluation Benchmarks [J]. Journal of the Association for Information Science and Technology, 2015, 66(12): 2626-2642.

[12] Phan N, Bailey P, Wilkinson R. Understanding the Relationship of Information Need Specificity to Search Query Length [C]. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07). New York: ACM, 2007: 709-710.

[13] Mu X, Lu K. Improving UMLS Metathesaurus Query Expansion Based on the Query Specificity and Length [C]. In: Proceedings of the ACM SIGKDD Workshop on Health Informatics.2012.

[14] Heine M H. An Investigation of the Relative Influences of Database Informativeness, Query Size and Query Term Specificity on the Effectiveness of Medline Searching [J]. Journal of Information Science, 1995, 21(3): 173-185.

[15] Ingwersen P, Jarvelin K. The Turn: Integration of Information Seeking and Retrieval in Context [M]. Springer, 2005.

[16] Ramírez G, de Vries A P. Relevant Contextual Features in XML Retrieval [C]. In: Proceedings of the 1st International Conference on Information Interaction in Context. New York: ACM, 2006: 56-65.

[17] Carletta J. Assessing Agreement on Classification Tasks: The Kappa Statistic [J]. Computational Linguistics, 1996, 22(2): 249-254.

[18] Siegel S, Castellan N J. Non-parametric Statistics for the Behavioral Sciences [J]. American Catholic Sociological Review, 1957, 18(2). DOI: 10.2307/3708383.

[19] Ponte J M, Croft W B. A Language Modeling Approach to Information Retrieval [C]. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1998: 275-281.

[20] Robertson S E, Jones K S. Relevance Weighting of Search Terms [J]. Journal of the American Society for Information Science, 1976, 27(3): 129-146.

作者贡献声明:

任珂: 文献调研, 数据处理及数据分析, 论文起草、修改及最终版本修订;
陆伟: 提出研究思路, 论文修改及最终版本修订;
丁恒: 数据分析, 论文修改。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。

[1] 任珂, 陆伟, 丁恒. eval-adhoc-dir.xlsx. 语言模型狄利克雷平滑在不同指标下的检索结果.

[2] 任珂, 陆伟, 丁恒. eval-adhoc-jm.xlsx. 语言模型线性插值平滑在不同指标下的检索结果.

[3] 任珂, 陆伟, 丁恒. eval-adhoc-bm25.xlsx. BM25 模型在不同指标下的检索结果.

[4] 任珂, 陆伟, 丁恒. 数据标注结果集.xlsx. 数据标注结果集.

收稿日期: 2016-07-18
收修改稿日期: 2016-09-01

The Impacts of Query Specificity on Information Retrieval

Ren Ke[1] Lu Wei[1, 2] Ding Heng[1]
[1](School of Information Management, Wuhan University, Wuhan 430072, China)
[2](Center for the Study of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: [**Objective**] This paper analyzes the impacts of query specificity on the effectiveness of information retrieval systems, aiming to improve the performance of search engine and user experience. [**Methods**] First, we manually constructed a labeling set for queries from the TREC Web Track. Second, we adopted the Dirichlet language model, linear interpolation language model and BM25 model to examine each query's performance. Finally, we used the average information retrieval evaluation index as the benchmark to explore the impacts of query specificity. [**Results**] For the highest-ranked results, the queries with narrower specificity had better retrieval performance than their boarder counterparts. [**Limitations**] The proposed method was only examined with data provided by TREC. More studies were needed to evaluate its performance with other data sets. [**Conclusions**] Search engines should focus on the precision of the highest ranked results, and then modify their retrieval model accordingly.

Keywords: Query intention Query specificity Retrieval result

ProQuest 发布 2016 年图书馆空间再利用调查的结果

最新发布的 ProQuest 对 600 多家图书馆的调查发现, 82%的学术图书馆认为空间再利用已经是或者即将是图书馆中最重要的事情之一。超过三分之一的学术图书馆对图书馆空间再利用的考虑已长达 5 年之久。

调查显示了图书馆空间再利用的各种创新方式, 25%的学术图书馆表示他们正在设立制造者空间和黑客空间。调查白皮书"用户需求不断进化下的学术图书馆空间进化"可在 http://bit.ly/libspacereclaim 通过注册免费获取。

ProQuest 北美区销售副总裁 Kevin Stehr 说: "尽管空间再利用计划如此受欢迎, 许多图书馆在进行空间再利用时仍然需要努力解决一些常见的问题。比如说, 只有 6%的图书馆为电子书分配了新的预算。"

"即使是那些重新分配印刷本预算的图书馆", Stehr 补充说, "也必须在科研人员对合作空间的渴望和他们对印刷本资料的喜爱之间进行平衡。" 正如一位受调查的图书馆员所说, "做好这一平衡需要深入了解所有格式资源的使用情况, 以便做出最好的馆藏分配决定。"

利用各种 ProQuest 解决方案和服务能帮助图书馆实现这种平衡并重新利用空间。比如, 该公司的标题快速匹配(Title Matching Fast)服务能帮助图书馆评估他们的印刷图书、数字图书和期刊杂志, 以做出数据驱动的馆藏建设方案; 该公司的数字存档和访问计划(Digital Archive and Access Program)能帮助图书馆将印刷图书馆藏转换为在线数字馆藏, 节省大量的空间, 同时又保持了其馆藏覆盖范围。

(编译自: http://www.proquest.com/about/news/2016/ProQuest-Releases-Results-of-Its-2016-Space-Reclamation-Survey.html)

(本刊讯)